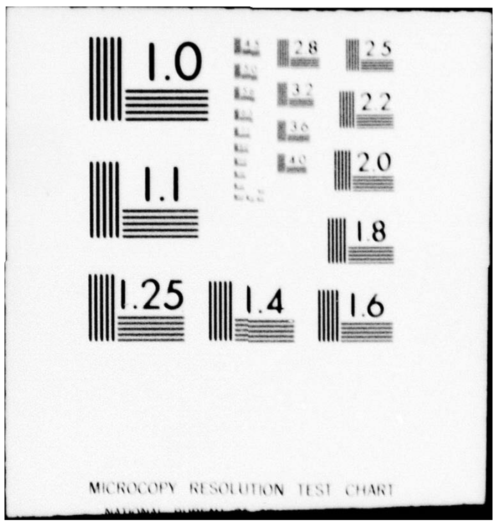
AD-A060 980 PURDUE UNIV LAFAYETTE IND DEPT OF STATISTICS F/G 12/1
ON SELECTION PROCEDURES FOR NORMAL POPULATIONS WITH COMMON KNOW--ETC(U)
JUL 78 S S GUPTA, A K SINGH N00014-75-C-0455
UNCLASSIFIED MIMEOGRAPH SER-78-14 NL

| OF |

AD
A060 980

END
DATE
FILMED
2-79
DDC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

6

On Selection Procedures for Normal Populations with Common known Coefficient of Variation with an Application to Multivariate Normal Populations*

by

10

Shanti S. Gupta and A. K. Singh
Purdue University

9 Technical rept.

Department of Statistics
Division of Mathematical Sciences

14 Mimeograph Series #78-14

11 July 1978

12 27p.

*This research was supported by the Office of Naval Research contract 15 N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

1B

On Selection Procedure for Normal Populations with Common known Coefficient of Variation with an Application to Multivariate Normal Populations*

by

Shanti S. Gupta and A. K. Singh
Purdue University

1. Introduction.

There are instances in the biological and physical sciences in which a linear relationship between the sample means and the sample standard deviations seems to exist. It is not unreasonable in such cases to assume that the population standard deviation σ is proportional to the population mean θ, i.e. $\sigma = b^{1/2}\theta$ where b is a positive constant. The unknown mean θ is assumed to be positive. Khan (1968) has considered the problem of estimation of θ for the normal distribution $N(\theta, b\theta^2)$ and has derived an unbiased estimator which has the minimum variance among a subclass of the class of unbiased estimators. Glesser and Healy (1975) have derived an estimator of θ which has the minimum mean square error among a larger class than that considered by Khan, thus proving the inadmissibility with respect to the squared error loss of Khan's estimator. However, the two estimators mentioned above are asymptotically equivalent, both being best asymptotically normal (BAN). The purpose of this paper is to investigate some problems of selection when the underlying populations are normal with unknown positive means $\theta_1, \ldots, \theta_k$ and a common known coefficient of variation $b^{1/2}$.

In Section 2 three procedures for selecting the normal population associated with the largest mean are proposed and investigated. In Section 3 pairwise asymptotic comparisons are made between these rules. In Section 4 some operating characteristics of the rule R_1 and R_3 are computed for a slippage type configuration of the normal means. Section 5 consists of an

---

*This research was supported in the Office of Naval Research contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

illustration of the use of the proposed rules. In Section 6, an application of the rules in a problem of selection for multivariate normal population is given.

2. Procedures for Selecting the Largest Normal Mean.

Let $\pi_1,\ldots,\pi_k$ be k ($\geq$ 2) independent univariate normal populations with means $\theta_1,\ldots,\theta_k$, respectively, and a common known coefficient of variation $b^{1/2}$. The goal is to select, on the basis of a sample $x_{i1},\ldots,x_{in}$ from π_i for each i, a subset which contains the population corresponding to the largest θ_i with probability at least P*, where P* is a preassigned constant (1/k < P* < 1). For each i = 1,...,k let

$$\bar{X}_i = \sum_{j=1}^{n} X_{ij}/n$$

and

$$S_i^2 = \sum_{j=1}^{n} (X_{ij}-\bar{X})^2/n.$$

Also let $B_n = (2b)^{-1/2}\Gamma((n-1)/2)/\Gamma(n/2)$. Then from Khan (1968) we have

1. $T_{1i} = \bar{X}_i$ and $T_{2i} = B_n n^{1/2} S_i$ are both unbiased estimators of θ_i.
2. The estimator

$$T_{3i} = \alpha^* T_{2i} + (1-\alpha^*)T_{1i} \tag{2.1}$$

where $\alpha^* = b/[b + n\{(n-1)\Gamma^2((n-1)/2)/2\Gamma^2(n/2) - 1\}]$

is the uniformly minimum variance unbiased estimator (UMVUE) among all estimators of the type $\alpha T_{2i} + (1-\alpha)T_{1i}$, $0 \leq \alpha \leq 1$, and is a BAN estimator with asymptotic variance $b\theta^2/[n(1+2b)]$. In fact, T_{3i} is the UMVUE among all estimators of the type $\alpha T_{2i} + (1-\alpha)T_{1i}$, $-\infty < \alpha < \infty$ [see Gleser and Healy (1975)].

For the problem of selection of a subset containing the largest mean the following rules will be investigated:

$$R_\ell: \text{ Select } \pi_i \text{ iff } T_{\ell i} \geq \sqrt{c_\ell} \max_{1\leq j\leq k} T_{\ell j},\ \ell = 1,2,3, \tag{2.2}$$

where c_ℓ $(0 < c_\ell < 1)$ is the smallest constant satisfying the basic P*-condition.

Probability of a Correct Selection.

Let $Y_{\ell i} = T^2_{\ell i}$ and $Y_{\ell(i)}$ represent the (unknown) Y-value associated with i-th largest mean $\theta_{[i]}$. Then

$$P(CS|R_\ell) = P(\max_{1\leq j\leq k-1} Y_{\ell(j)} \leq \frac{1}{c_\ell} Y_{\ell(k)}). \tag{2.3}$$

Now, since $(n/b\theta^2_{[j]})_{1(j)}$ has a non-central $\chi^2(1,\lambda)$ distribution with 1 degree of freedom and non-centrality paramenter $\lambda = n/b$, we have

$$P(CS|R_1) = \int_0^\infty \prod_{j=1}^{k-1} G_\lambda(u\theta^2_{[k]}/c_1\theta^2_{[j]})dG_\lambda(u) \tag{2.4}$$

where $G_\lambda(\cdot)$ is the cumulative distribution function of a $\chi^2(1,\lambda)$ distribution. The integrand in (2.4) is clearly an increasing function of $\theta_{[k]}/\theta_{[j]}$ and hence the infimum of $P(CS|R_1)$ is attained at $\theta_1 = \ldots = \theta_k$. Hence the equation for c_1 is

$$\int_0^\infty G^{k-1}_\lambda(u/c_1)dG_\lambda(u) = P^*. \tag{2.5}$$

Now, from Han (1975), we have

$$G_\lambda(y) = \Phi(\sqrt{\lambda} + \sqrt{y}) - \Phi(\sqrt{\lambda} - \sqrt{y}). \tag{2.6}$$

It follows from (2.5) and (2.6) that c_1 can be determined from

$$\frac{1}{2}\int_0^\infty [\Phi(\sqrt{n/b}+\sqrt{u/c_1})-\Phi(\sqrt{n/b}-\sqrt{u/c_1})]^{k-1}$$

$$\cdot\frac{1}{\sqrt{u}}[\phi(\sqrt{n/b}+\sqrt{u})+\phi(\sqrt{n/b}-\sqrt{u})]du = P^*. \tag{2.7}$$

For selected values of k, P* and $\sqrt{\lambda}=\sqrt{n/b}$, values of c_1 satisfying (2.7) have been computed, and tables of these values are given at the end of this paper.

Similarly, using the fact that $n\,Y_{2(j)}/\,b\,B_n^2\theta_{[j]}^2$ has a central chi-square distribution with n - 1 degrees of freedom (df) it can be seen that the smallest c_2 is given by

$$\int_0^\infty [G_{n-1}(u/c_2)]^{k-1}\, dG_{n-1}(u) = P^*, \tag{2.8}$$

where G_{n-1} denotes the distribution function of a central chi-square random variable with df n-1.

The constant c_3 for the rule R_3 is obtained by using the asymptotic distribution of T_{3i}. The equation in this case is the same as (2.7) with n/b replaced by n(1+2b)/b.

Expected Size of the Selected Subset.

The size of the subset selected by a rule of the form (2.1) is a random variable and its expected value is used as a criteria of efficiency of a selection procedure [see, for example, Gupta (1965)]. For a rule R, let $P_{\underline{\theta}}(i|R) \equiv P_{\underline{\theta}}(i,k,n,P^*|R)$ denote the probability of selecting the population associated with $\theta_{[i]}$ when $\underline{\theta}$ is the true parameter value. Then

$$E_{\underline{\theta}}(S|R) = \sum_{i=1}^{k} P_{\underline{\theta}}(i,k,n,P^*|R). \tag{2.9}$$

Now, for the rule R_2, we have

$$P_{\underline{\theta}}(i,k,n,P^*|R_2) = \int_0^\infty \prod_{\substack{j=1\\j\neq i}}^{k} [G_{n-1}(u\theta^2_{[i]}c_2\theta^2_{[j]})]dG_{n-1}(u), \tag{2.10}$$

and for $\ell = 1,3$

$$P_{\underline{\theta}}(i,k,n,P^*|R_\ell) = \int_0^\infty \prod_{\substack{j=1\\j\neq i}}^{k} [G_\lambda(u\theta^2_{[i]}|c_\ell\theta^2_{[j]})]dG_\lambda(u) \tag{2.11}$$

where $\lambda = n/b$ for R_1, and $= n(1+2b)/b$ for R_3.

Expressions for $E_{\underline{\theta}}(S|R)$ for the rules under consideration can be obtained from (2.9), (2.10) and (2.11).

Remark.

It is clear from (2.10) and (2.11) that the selection rules given by (2.2) are strongly monotone [see Santner (1975)], i.e.

$P_{\underline{\theta}}(i,k,n,P^*|R_\ell)$ is $\uparrow$ $\theta_{[i]}$ if remaining components of $\underline{\theta}$ are kept fixed

$\downarrow$ $\theta_{[j]}$, $j \neq i$ if remaining components of $\underline{\theta}$ are kept fixed.

Expected Sum of Ranks.

If the population associated with mean $\theta_{[i]}$ is given rank i, then the expected sum of ranks of the populations selected by a rule R is given by

$$\Psi_{\underline{\theta}}(k,P^*,n|R) = \sum_{i=1}^{k} i\, P(i,k,P^*,n|R)/k = k\Psi_1(k,P^*,n|R). \tag{2.12}$$

Expected sum of ranks by itself does not tell much, and it is more meaningful to look at the ratio $\Psi_{\underline{\theta}}/E_{\underline{\theta}}$.

3. Asymptotic Relative Efficiencies of R_1 and R_2 with Respect to R_3.

In this section on asymptotic comparison is made among the three rules when $k = 2$ and the parameters are in a slippage-type configuration. Suppose we are given two normal populations $N(\theta, b\theta^2)$ and $N(\Delta\theta, b\Delta^2\theta^2)$ with $\theta > 0$, $b > 0$ and $\Delta > 1$. The population with mean θ will be referred to as the non-best population.

Let S^* denote the number of non-best populations selected by a rule. Then small values of S^* are desirable and therefore, consistent with the basic P^*-condition, one would like to keep the expected value of S^* as small as possible. For a given ε $(0 < \varepsilon < 1)$, let $N_R(\varepsilon)$ be the number of observations needed so that

$$E(S^*|R) = \varepsilon . \tag{3.1}$$

The following definition of ARE will be used [see Barlow and Gupta (1969)]:

Definition 3.1: The ARE of a rule R_1 relative to another rule R is given by

$$\mathrm{ARE}(R_1, R; \theta) = \lim_{\varepsilon \downarrow 0} \frac{N_R(\varepsilon)}{N_{R_1}(\varepsilon)} .$$

Consider first the rules R_1 and R_3. For the sake of convenience we will suppress the subscript ℓ $(\ell = 1,3)$. We have

$$P(CS|R) = P(W \leq \Delta^2/c) \tag{3.2}$$

where $W = \Delta^2 Y_{(1)}/Y_{(2)}$ is approximately distributed as an F random variable with degrees of freedom (m,m) where $m = (1+\lambda)^2/(1+2\lambda)$. Hence the smallest

c satisfying the basic P*-condition is given by

$$F_{m,m}(c^{-1}) = P^*. \tag{3.3}$$

It follows from Zelen and Severo (1964) that

$$c \approx e^{-2w}, \tag{3.4}$$

where

$$w = y(h+\mu)^{1/2}/h$$

$$y = \Phi^{-1}(P^*)$$

$$h = \lambda^2/(1+2\lambda) \tag{3.5}$$

$$\mu = (y^2-3)/6\,.$$

Next, we have

$$E(S^*|R) = P(W \le 1/(c\Delta^2)) = F_{m,m}(1/(c\Delta^2)). \tag{3.6}$$

Equating (3.6) to ε, we have

$$c\Delta^2 \approx e^{-2w'}, \tag{3.7}$$

where w' is same as w of (3.5) with y replaced by $y' = \Phi^{-1}(\varepsilon)$. Hence

$$w - w' = \log \Delta. \tag{3.8}$$

For $R = R_1$ we have $\lambda = n/b$ and therefore, for n sufficiently large, (3.8) leads to

$$N_{R_1}(\varepsilon) = 2b(y-y')^2/(\log \Delta)^2. \tag{3.9}$$

Similarly

$$N_{R_3}(\varepsilon) = 2b(y-y')^2/(1+2b)\log\Delta)^2, \tag{3.10}$$

and therefore

$$\mathrm{ARE}(R_1,R_3) = \lim_{\varepsilon\downarrow 0} \frac{N_{R_3}(\varepsilon)}{N_{R_1}(\varepsilon)} = (1+2b)^{-1}. \tag{3.11}$$

In a similar way we can show that

$$N_{R_2}(\varepsilon) = ((y'-y)/\log\Delta)^2 + 2, \tag{3.12}$$

which, together with (3.10), gives

$$\mathrm{ARE}(R_2,R_3) = \lim_{\varepsilon\downarrow 0} \frac{\frac{2b}{1+2b}\left(\frac{y-y'}{\log\Delta}\right)^2}{\left(\frac{y-y'}{\log\Delta}\right)^2 + 2} \tag{3.13}$$

$$= 2b/(1+2b),$$

since $y' \to -\infty$ as $\varepsilon \to 0$.

It is clear from (3.11) and (3.13) that R_1 (R_2) and R_3 are asymptotically equivalent if b is small (large) and therefore, for reasons of being simpler to use, may be preferred over R_3 for b small (large). R_3, however, does provide considerable saving in terms of sample size for intermediate values of b. This fact is also intuitively appealing, as R_3 uses more information contained in the sample than either of the other two rules.

The above discussion seems to indicate that R_3 should be preferred over R_1 and R_2 if b is of moderate size. In the next section we investigate the rules R_1 and R_3 further by numerically computing the values of the functions $P_{\underline{\theta}}$, $E_{\underline{\theta}}/k$, $\Psi_{1\underline{\theta}}$ and $\Psi_{\underline{\theta}}/E_{\underline{\theta}}$ of Section 2 associated with these rules.

4. On the Performance of the Rule R_1 and R_3.

Let the k population means be $\theta, \Delta\theta, \ldots, \Delta^{k-1}$ where $\Delta > 1$ is a known constant. Then $\theta_{[i]}/\theta_{[j]} = \Delta^{i-j}$ and

$$P(i,k,n,P^*|R) = \int_0^\infty \left(\prod_{\substack{j=1 \\ j\neq i}}^{k} \Phi[\sqrt{\lambda} + \Delta^{i-j}\sqrt{u/c_3}) - \Phi(\sqrt{\lambda} - \Delta^{i-j}\sqrt{u/c_3})]\right) \tag{4.1}$$

$$\cdot \frac{1}{2\sqrt{u}} [\phi(\sqrt{\lambda} + \sqrt{u}) + \phi(\sqrt{\lambda} - \sqrt{u})]du .$$

where $\lambda = n/b$ for $R = R_1$, and $\lambda = n(1+2b)/b$ for $R = R_3$.

The expected subset size $E_{\underline{\theta}}(S|R)$, the expected average rank $\Psi_{1\underline{\theta}}(S|R)$ and the ratio $\Psi_{\underline{\theta}}/E_{\underline{\theta}}$ can now be obtained from (2.8) and (2.11).

For selected values of k, P^*, Δ and $\lambda^{1/2}$, the probability given by (4.1) has been computed and tables are given at the end of this paper. For example, if $k = 4$, $P^* = .90$, $\Delta = 3.0$, $\lambda^{1/2} = 1.5$, then R selects the second best and third best populations with probabilities .788 and .345, respectively. The probability of selecting the best population has to be at least .90, and actually equals .972 in this case. Tables are also provided for the expected proportion of populations in the subset $(E_{\underline{\theta}}(S|R_3)/k)$, the expected average rank of the selected subset, and the ratio of expected sum of ranks to theexpected subset size. For example, if $k = 4$, $P^* = .90$, $\Delta = 3.0$, $\lambda^{1/2} = 1.5$, then the expected sum of ranks is 6.99 which compares with 7, the sum of ranks of the two best populations since the expected subset size is 2.16.

<u>Remarks</u>: It is clear from (4.1) that

1. For fixed k, λ, P^* and $i = 1(k)$, the probability fo selecting the i-th population strictly decreases (increases) to 0(1) as Δ is increased.

2. $$\psi > \frac{(k+1)}{2} \int_0^\infty \left(\prod_{j=2}^{k} [\Phi(\sqrt{\lambda} + \Delta^{1-j}\sqrt{u/c_3}) - \Phi(\sqrt{\lambda} - \Delta^{1-j}\sqrt{u/c_3})] \right)$$

$$\cdot \frac{1}{2\sqrt{u}} [\phi(\sqrt{\lambda} + \sqrt{u}) + \phi(\sqrt{\lambda} - \sqrt{u})]du$$

$$> \frac{(k+1)}{2} \int_0^\infty [\Phi(\sqrt{\lambda} + \Delta^{1-k}\sqrt{u/c_3}) - \Phi(\sqrt{\lambda} - \Delta^{1-k}\sqrt{u/c_3})]^{k-1}$$

$$\cdot \frac{1}{2\sqrt{u}} [\phi(\sqrt{\lambda} + \sqrt{u}) + \phi(\sqrt{\lambda} - \sqrt{u})]du .$$

5. An Example

In this section we illustrate the use of the proposed selection rules. Using the IMSL subroutine GGNOR, 8 pseudo-random numbers were generated from each of 5 normal populations with means $\theta_1 = 1.0$, $\theta_2 = 2.5$, $\theta_3 = .25$, $\theta_4 = 4.5$, $\theta_5 = 1.5$ and common coefficient of variation $b = .5$. Values of the statistics T_{1i}, T_{2i} and T_{3i}, $i = 1,\ldots,5$ of Section 2 computed from these samples are given below:

j	1	2	3	4	5
T_{1j}	.805	2.417	.217	5.396	1.015
T_{2j}	.602	2.648	.251	3.787	1.554
T_{3j}	.712	2.523	.232	4.661	1.261

For the rule R_1 we have $\sqrt{\lambda} = 4.0$. For $P^* = .95$, the tabulated value of the constant c_1 is .2038. Hence R_1 reduces to

$$R_1: \text{ select } \pi_i \text{ iff } T_{1i} \geq 5.396\sqrt{.2038} = 2.436 .$$

Thus using R_1 we select the best population π_4.

Similarly, for $P^* = .95$, $c_3 = .2401$ and the rule R_3 reduces to

$$R_3: \text{ select } \pi_i \text{ iff } T_{3i} \geq 2.284,$$

resulting in the selection of π_2 and π_4.

For the rule R_2, the constant c_2 is .242 [see Table IB of Gupta (1963)], so that in this case we have

$$R_2: \text{ select } \pi_i \text{ iff } T_{2i} \geq 1.863,$$

which again results in the selection of π_2 and π_4.

6. Selection of the Largest Characteristic Root of a Multivariate Population.

Suppose that n independent observation vectors $\underline{x}_1,\ldots,\underline{x}_n$ are recorded on a k-dimensional random vector which is distributed according to the multivariate normal law with mean vector $\underline{\mu}$ and non-singular covariance matrix Σ. Assume that Σ has distinct characteristic roots $0 < \lambda_1 < \ldots < \lambda_k$. The *largest characteristic root* λ_k *is of interest in many practical situations* and one may wish to select a subset of $\{\lambda_i: \ i = 1,\ldots,k\}$ which contains λ_k.

Let S denote the sample covariance matrix

$$S = \frac{1}{n-1} \sum_{j=1}^{n} (\underline{x}_j - \bar{\underline{x}})(\underline{x}_j - \bar{\underline{x}})'$$

where $\bar{\underline{x}}$ is the vector of the sample means. Let $0 < \ell_1 < \ldots < \ell_k$ denote the characteristic roots of S. It has been shown by Anderson (1951) that, for large n, $\sqrt{n/2}(\ell_i-\lambda_i)/\lambda_i$ is approximately a standard normal variate and is

independent of λ_j, $j \neq i$. It follows that the problem of selection of the largest root λ_k is equivalent to that considered in Section 2 with $\theta_i = \sqrt{n}\,\lambda_i$ and $b = 2/n$. Moreover, since b is small in this case, it follows from (3.11) that the rule R_1 may be preferred to R_2 or R_3 for reasons of simplicity.

Table showing the values of $c = c(k,P^*,\sqrt{\lambda})$ for the rules R_1 and R_3.

$P^* = .90$

k \ $\sqrt{\lambda}$	1.5	2.0	2.5	3.0	4.0	5.6789
2	.0533	.1144	.1943	.2691	.3873	.4536
3	.0311	.0764	.1429	.2090	.3203	.3986
4	.0230	.0634	.1234	.1847	.2913	.3729
5	.0191	.0566	.1125	.1751	.2740	.3570
8	.0143	.0465	.0959	.1491	.2462	.3304
11	.0123	.0418	.0876	.1381	.2318	.3159
14	.0111	.0388	.0823	.1310	.2223	.3063

For given k and n, the entries in this table are the values of c which satisfy

$$\int_0^\infty G_\lambda^{k-1}(u/c)\,dG_\lambda(u) = P^*$$

where $G(\cdot)$ is the cdf of non-central $\chi^2(1,\lambda)$, and $\lambda = \begin{cases} n/b & \text{for } R_1 \\ n(1+2b)/b & \text{for } R_3 \end{cases}$.

Table showing the values of $c = c(k, P^*, \sqrt{\lambda})$ for the rules R_1 and R_3.

$P^* = .95$

k \ $\sqrt{\lambda}$	1.5	2.0	2.5	3.0	4.0	5.6789
2	.0170	.0407	.0984	.1664	.2845	.2995
3	.0123	.0247	.0717	.1296	.2372	.2662
4	.0106	.0185	.0612	.1146	.2164	.2502
5	.0100	.0155	.0557	.1060	.2038	.2401
8	.0083	.0116	.0466	.0927	.1836	.2231
11	.0076	.0100	.0419	.0860	.1730	.2137
14	.0071	.0091	.0389	.0819	.1661	.2074

For given k and n, the entries in this table are the values of c which satisfy

$$\int_0^\infty G_\lambda^{k-1}(u/c)\,dG_\lambda(u) = P^*$$

where $G(\cdot)$ is the cdf of non-central $\chi^2(1,\lambda)$, and $\lambda = \begin{cases} n/b & \text{for } R_1 \\ n(1+2b)/b & \text{for } R_3 \end{cases}$.

For the rules R_1 and R_3, and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$ this tables gives the probability of selecting the normal population with rank i when the population with mean $\Delta^{i-1}\theta$ has rank i, $i = 1, \ldots, k$; the constant $\Delta > 1$ and the common coefficient of variation b are known.

$P^* = .90$, $\sqrt{\lambda} = 3.0$

k	i \ Δ	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
2	1	.700	.469	.293	.183	.119	.080	.057	.043
	2	.968	.984	.990	.993	.995	.995	.996	.997
3	1	.440	.093	.019	.005	.002	.001	.000	.000
	2	.780	.575	.389	.256	.170	.116	.083	.061
	3	.976	.988	.992	.994	.995	.996	.997	.997
4	1	.162	.004	.000	.000	.000	.000	.000	.000
	2	.497	.117	.025	.007	.002	.001	.001	.000
	3	.814	.625	.440	.297	.201	.139	.099	.073
	4	.979	.989	.993	.995	.996	.996	.997	.998
5	1	.027	.000	.000	.000	.000	.000	.000	.000
	2	.179	.005	.000	.000	.000	.000	.000	.000
	3	.521	.128	.028	.008	.003	.001	.001	.000
	4	.827	.646	.462	.317	.216	.150	.107	.079
	5	.981	.990	.993	.995	.996	.996	.997	.998

Here $\lambda = \begin{cases} n/b & \text{for } R_1 \\ n(1+2b)/b & \text{for } R_3 \end{cases}$.

For the rules R_1 and R_3, and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$, this tables gives the expected average rank of the selected subset (top), the expected proportion of the populations selected in the subset (middle) and the ratio of expected sum of ranks to expected subset size when the population with mean $\Delta^{i-1}\theta$ has rank i; the constant $\Delta > 1$ and the common coefficient of variance b are known.

$P^* = .90,\ \sqrt{\lambda} = 3.0$

k \ Δ	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
2	1.318	1.219	1.137	1.085	1.054	1.036	1.025	1.018
	.834	.727	.642	.588	.557	.538	.527	.520
	1.580	1.677	1.771	1.845	1.892	1.926	1.945	1.958
3	1.643	1.402	1.258	1.167	1.109	1.074	1.052	1.038
	.732	.552	.467	.418	.389	.371	.360	.353
	2.245	2.540	2.694	2.792	2.851	2.895	2.922	2.941
4	1.879	1.518	1.335	1.221	1.148	1.101	1.072	1.053
	.613	.434	.365	.325	.300	.284	.274	.268
	3.065	3.498	3.658	3.757	3.827	3.877	3.912	3.929
5	2.032	1.586	1.380	1.253	1.170	1.117	1.083	1.061
	.507	.354	.297	.264	.243	.230	.221	.215
	4.008	4.480	4.646	4.746	4.815	4.857	4.900	4.935

Here $\lambda = \begin{cases} n/b & \\ n(1+2b)/b & \text{for } R_3 \end{cases}$.

For the rules R_1 and R_3, and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$, this table gives probability of selecting the normal population with rank i when the population with mean $\Delta^{i-1}\theta$ has rank i, $i = 1, \ldots, k$; the constant $\Delta > 1$ and the common coefficient of variation b are known.

$P^* = .90$, $\sqrt{\lambda} = 4.0$

k	i \ Δ	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
2	1	.577	.270	.114	.051	.025	.013	.007	.005
	2	.985	.996	.998	.999	.999	.999	1.000	1.000
3	1	.219	.013	.001	.000	.000	.000	.000	.000
	2	.677	.363	.169	.078	.038	.020	.012	.007
	3	.990	.997	.999	.999	.999	.999	1.000	1.000
4	1	.029	.000	.000	.000	.000	.000	.000	.000
	2	.264	.017	.001	.000	.000	.000	.000	.000
	3	.723	.414	.203	.096	.048	.026	.015	.009
	4	.992	.997	.999	.999	.999	1.000	1.000	1.000
5	1	.001	.000	.000	.000	.000	.000	.000	.000
	2	.036	.000	.000	.000	.000	.000	.000	.000
	3	.296	.021	.002	.000	.000	.000	.000	.000
	4	.750	.448	.227	.110	.055	.030	.017	.010
	5	.993	.998	.999	.999	.999	1.000	1.000	1.000

Here $\lambda = \begin{cases} n/b & \text{for } R_1 \\ n(1+2b)/b & \text{for } R_3 \end{cases}$.

For the rules R_1 and R_3, and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$, this tables gives the expected average rank of the selected subset (top), the expected proportion of the populations selected in the subset (middle) and the ratio of expected sum of ranks to expected subset size when the population with mean $\Delta^{i-1}\theta$ has rank i; the constant $\Delta > 1$ and the common coefficient of variance b are known.

$P^* = .90,\ \sqrt{\lambda} = 4.0$

k \ Δ	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
2	1.273	1.130	1.055	1.024	1.012	1.006	1.003	1.002
	.781	.633	.556	.525	.512	.506	.504	.502
	1.630	1.785	1.897	1.950	1.977	1.988	1.990	1.996
3	1.514	1.243	1.111	1.051	1.025	1.013	1.008	1.005
	.629	.458	.389	.359	.346	.340	.337	.336
	2.407	2.714	2.856	2.928	2.962	2.979	2.991	2.991
4	1.673	1.317	1.151	1.072	1.035	1.019	1.011	1.006
	.502	.357	.301	.274	.262	.256	.254	.252
	3.333	3.689	3.824	3.912	3.950	3.980	3.980	3.992
5	1.785	1.369	1.181	1.087	1.044	1.023	1.013	1.008
	.415	.293	.245	.222	.211	.206	.203	.202
	4.301	4.672	4.820	4.896	4.948	4.966	4.990	4.990

Here $\lambda = \begin{cases} n/b & \text{for } R_1 \\ n(1+2b)/b & \text{for } R_3 \end{cases}$.

For the rules R_1 and R_3, and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$, this tables gives probability of selecting the normal population with rank i when the population with mean $\Delta^{i-1}\theta$ has rank i, $i = 1, \ldots, k$; the constant $\Delta > 1$ and the common coefficient of variation b are known.

$P^* = .95, \sqrt{\lambda} = 3.0$

k	i \ Δ	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
2	1	.840	.666	.483	.336	.231	.161	.115	.085
	2	.983	.991	.994	.995	.996	.997	.997	.998
3	1	.651	.208	.051	.015	.005	.002	.001	.001
	2	.887	.752	.588	.435	.314	.226	.164	.122
	3	.986	.992	.995	.996	.997	.997	.998	.999
4	1	.343	.016	.001	.000	.000	.000	.000	.000
	2	.698	.249	.064	.019	.007	.003	.001	.001
	3	.906	.789	.638	.487	.360	.264	.194	.145
	4	.988	.993	.995	.996	.997	.998	.998	.999
5	1	.096	.000	.000	.000	.000	.000	.000	.000
	2	.378	.019	.001	.000	.000	.000	.000	.000
	3	.726	.276	.074	.022	.008	.003	.002	.001
	4	.916	.810	.667	.520	.391	.290	.216	.162
	5	.989	.994	.995	.996	.997	.998	.999	.999

Here $\lambda = \begin{cases} n/b & \text{for } R_1 \\ n(1+2b)/b & \text{for } R_3 \end{cases}$.

For the rules R_1 and R_3, and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$, this tables gives the expected average rank of the selected subset (top), the expected proportion of the populations selected in the subset (middle) and the ratio of expected sum of ranks to expected subset size when the population with mean $\Delta^{i-1}\theta$ has rank i; the constant $\Delta > 1$ and the common coefficient of variance b are known.

$P^* = .95$, $\sqrt{\lambda} = 3.0$

k \ Δ	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
2	1.402	1.323	1.235	1.163	1.111	1.077	1.055	1.040
	.911	.828	.739	.665	.613	.579	.556	.541
	1.539	1.598	1.671	1.749	1.812	1.860	1.897	1.922
3	1.795	1.563	1.404	1.291	1.208	1.149	1.108	1.080
	.841	.651	.545	.482	.439	.409	.388	.374
	2.134	2.401	2.576	2.678	2.752	2.809	2.856	2.888
4	2.102	1.713	1.506	1.371	1.271	1.197	1.145	1.108
	.734	.512	.425	.376	.341	.316	.299	.286
	2.864	3.346	3.544	3.646	3.727	3.788	3.829	3.874
5	2.328	1.815	1.574	1.426	1.315	1.232	1.172	1.129
	.621	.420	.348	.308	.279	.258	.243	.232
	3.749	4.321	4.523	4.630	4.713	4.775	4.823	4.866

Here $\lambda = \begin{cases} n/b & \text{for } R_1 \\ n(1+2b)/b & \text{for } R_3 \end{cases}$.

For the rules R_1 and R_3, and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$, this tables gives probability of selecting the normal population with rank i when the population with mean $\Delta^{i-1}\theta$ has rank i, i = 1,...,k; the constant $\Delta > 1$ and the common coefficient of variation b are known.

$P^* = .95, \sqrt{\lambda} = 4.0$

k	i \ Δ	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
2	1	.734	.427	.212	.101	.051	.027	.016	.010
	2	.992	.998	.999	.999	.999	1.000	1.000	1.000
3	1	.375	.033	.003	.000	.000	.000	.000	.000
	2	.808	.529	.290	.149	.077	.042	.024	.015
	3	.995	.998	.999	.999	1.000	1.000	1.000	1.000
4	1	.077	.000	.000	.000	.000	.000	.000	.000
	2	.428	.043	.004	.000	.000	.000	.000	.000
	3	.839	.581	.335	.178	.094	.051	.030	.018
	4	.996	.998	.999	.999	1.000	1.000	1.000	1.000
5	1	.004	.000	.000	.000	.000	.000	.000	.000
	2	.091	.000	.000	.000	.000	.000	.000	.000
	3	.464	.051	.005	.001	.000	.000	.000	.000
	4	.857	.613	.366	.200	.107	.059	.034	.021
	5	.996	.999	.999	.999	1.000	1.000	1.000	1.000

Here $\lambda = \begin{cases} n/b & \text{for } R_1 \\ n(1+2b)/b & \text{for } R_3 \end{cases}$.

For the rules R_1 and R_3, and the configuration $(\theta, \Delta\theta, \ldots, \Delta^{k-1}\theta)$, this tables gives the expected average rank of the selected subset (top), the expected proportion of the populations selected in the subset (middle) and the ratio of expected sum of ranks to expected subset size when the population with mean $\Delta^{i-1}\theta$ has rank i; the constant $\Delta > 1$ and the common coefficient of variance b are known.

$P^* = .95$, $\sqrt{\lambda} = 4.0$

k \ Δ	1.50	2.0	2.5	3.0	3.5	4.0	4.5	5.0
2	1.359	1.211	1.105	1.050	1.025	1.013	1.007	1.004
	.863	.712	.605	.550	.525	.513	.508	.505
	1.575	1.701	1.826	1.909	1.952	1.975	1.982	1.988
3	1.658	1.362	1.193	1.099	1.051	1.027	1.016	1.009
	.726	.520	.431	.383	.359	.347	.341	.338
	2.284	2.619	2.768	2.869	2.928	2.960	2.979	2.985
4	1.858	1.455	1.252	1.133	1.070	1.038	1.022	1.013
	.585	.406	.335	.295	.273	.263	.257	.254
	3.176	3.584	3.737	3.841	3.919	3.947	3.977	3.988
5	1.998	1.520	1.295	1.160	1.085	1.047	1.027	1.016
	.483	.333	.274	.240	.221	.212	.207	.204
	4.137	4.565	4.726	4.833	4.910	4.939	4.961	4.980

Here $\lambda = \begin{cases} n/b & \text{for } R_1 \\ n(1+2b)/b & \text{for } R_3 \end{cases}$.

REFERENCES

Barlow, R. E. and Gupta, S. S. (1969). Selection procedures for restricted families of distributions. Ann. Math. Statist. 40, 905-917.

Gleser, L. J. and Healy, J. D. (1975). Estimating the mean of a normal distribution with known coefficient of variation. Mimeo. Series No. 407, Department of Statistics, Purdue University, West Lafayette, IN.

Gupta, S. S. (1963). On a selection and ranking procedure for gamma populations. Ann. Inst. Statist. Math. 14, 199-216.

Gupta, S. S. (1965). On some multiple decision (selection and ranking) rules. Technometrics 7, 225-245.

Han, C. P. (1975). Some relationships between noncentral chi-squared and normal distributions. Biometrika 62, 213-214.

Khan, R. (1968). A note on estimating the mean of a normal distribution with known coefficient of variation. J. Amer. Statist. Assoc. 63, 1039-1041.

Santner, T. J. (1975). A restricted subset selection approach to ranking and selection problems. Ann. Statist. 3, 334-349.

Zelen, M. and Severo, N. C. (1964). Probability Functions. Chapter 26 of "Handbook of Math. Functions". Edited by M. Abramowitz and I. Stegun. National Bureau of Standards.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Mimeograph Series No. 78-14	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On Selection Procedures for Normal Populations with Common known Coefficient of Variation with an Application to Multivariate Normal Populations		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Mimeo. Series No. 78-14
7. AUTHOR(s) Shanti S. Gupta and A. K. Singh		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Purdue University West Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-243
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		12. REPORT DATE July, 1978
		13. NUMBER OF PAGES 24
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release, distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Uniformly minimum variance, best asymptotically normal, asymptotic relative efficiency, characteristic roots		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) There are instances in the biological and physical sciences in which a linear relationship between the sample means and the sample standard deviations seems to exist. It is not unreasonable in such cases to assume that the population standard deviation σ is a multiple of the population mean θ, i.e. $\sigma = b^{1/2}\theta$, $b > 0$, and θ also is assumed to be positive. The purpose of this paper is to investigate some problems of selection when the k under-		

Theta

Next page

Sigma (sq. rt. b) Theta, Theta

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Theta sub1 Theta sub k

lying populations are normal with unknown positive means $\theta_1,\ldots,\theta_k$ and a common known coefficient of variation $b^{1/2}$. Three selection rules are investigated, and some comparisons are made among these rules.

Sq. rt. b